卷积神经网络在乐器板材优劣识别中的应用研究* *

黄英来，李晓霜，赵 鹏

(东北林业大学 信息与计算机工程学院, 哈尔滨 150040)

摘 要：目前民族乐器板材振动信号识别算法具有特征提取复杂且耗时长等缺点，针对此问题，提出了一种基于卷积神经网络的木材振动信号分类识别算法，实现了乐器板材优劣的判别。卷积神经网络将特征提取和分类过程结合来进行神经网络的训练，具有识别度高、鲁棒性好等优点。首先重点分析和讨论了提取木材振动信号的语谱图特征，然后应用卷积神经网络结合网格搜索的方法进行参数调优。为了防止过拟合，还应用了 ReLU 和 Dropout 等新技术，得到最终分类结果。实验证明，测试样本准确率达到 96%，明显优于传统方法。该方法可减小人工测量的误差，加快板材的选取时间，为民族乐器制造领域的选材提供了一种更加实用的方法。

关键词：卷积神经网络；网格搜索；语谱图；木材振动信号

中图分类号：TP391

Research on wood quality for musical instrument recognition using convolutional neural network

Huang Yinglai, Li Xiaoshuang, Zhao Peng

(*College of Information and Computer Engineering*, *Northeast Forestry University*, *Harbin* 150040, *China*)

Abstract: At present, the vibration signal recognition algorithm for national musical instrument plate has the shortcomings of complex feature extraction and time-consuming. To solve this problem, this paper proposed a classification algorithm of wood vibration signal based on convolution neural network, to identify the quality of the musical instrument. Convolution neural network combines feature extraction and classification process to train the neural network, which owns the advantages of high recognition rate and good robustness. Firstly, this paper mainly analyzed and discussed spectrogram characteristics of the extraction of wood vibration signals. Then combining convolution neural network and grid search method, it can adjust the parameters. In order to avoid over-fitting, the final classification results were obtained by using new technologies such as ReLU and Dropout. The experiments showed that the accuracy of the test sample reached 96%, which are obviously better than the traditional method. This method can reduce the error of manual measurement and speed up the selection time of the plate, and provide a more convenient method for the selection of the national musical instrument manufacturing field.

Key Words: convolutional neural network; grid search; spectrogram; wood vibration signal

0 引言

近些年来，随着我国经济的快速发展和我国人民物质生活的提高，人们开始喜欢上了民族乐器，民族乐器演奏出的优美的旋律不仅可以陶冶品格情操，拓宽艺术视野，还可以缓解高强度的工作压力，甚至有很多外国人也对民族乐器兴趣大增，这使得我国的民族乐器产业也有了一定程度的发展。中国民族乐器的发展历史悠久[1]，并且作为中国传统文化的载体，拥有丰富的文化底蕴和民族特色，乐器的构造和种类也越来越多样化，人类对于乐器各方面的要求也变的越来越高。许多民族乐器的制作都离不开木材，木材的声学振动特性也大大影响着乐器的质量，因此对乐器板材质量优劣识别的研究和应用具有重要的现实意义。

当前国内外常用的乐器板材质量优劣识别方法有：中国林业科学研究院木材工业研究所的李源哲、李先泽，与北京乐器研究所的汪溪泉，王书勤提出了声波激发试样振动的方法对我国 31 种主要树种进行声学性能研究[2]；东北林业大学的沈隽博士与刘一星教授全面、系统地研究了纤丝角云杉属木材的声振

基金项目：国家自然科学基金资助项目（31670717）；国家教育部新世纪优秀人才专项基金资助项目（NCET-12-0809）；中央高校基本科研业务费专项基金资助项目（2572014CB25）

作者简介：黄英来（1978-），男，内蒙古赤峰人，副教授，博士，主要研究方向为信号处理与计算机智能识别（nefuhyl@163.com）；李晓霜（1993-），女，辽宁锦州人，硕士研究生，主要研究方向为深度学习与声音识别；赵鹏（1972-），男，黑龙江阿城人，教授，博士，主要研究方向为图像处理、模式识别和光学测量.

动特性，分析了各声学参数的变异规律以及其相互间的联系，揭示了云杉属木材各种构造因子对振动性能产生的影响[3]。则元京通过测定针叶树材动态弹性模量 E'、挠性振动内磨擦 Q^{-1} 及静曲弹性模量 E 等参数来评定木材的声学性质[4]；Sobue、Tonosaki 等人对木材的动态弹性模量，辐射阻尼常数 R、声速、声音特性阻抗ω、动力损耗角正切 tan δ /E 及 tan δ 等参数来评价说明云杉属木材是制作乐器音板的最佳用材[5, 6]；Treu 等人采用弹性测试仪与应力波无损评价技术对乐器音板的云衫属木材进行分等[7]。虽然这些方法都实现了乐器板材质量优劣的选择，但是大多是从木材声学属性参数方面进行实验及分析的，并且过程复杂且耗时较长，这些都是传统方法的局限性。目前随着深度学习和计算机技术以及声音识别技术的发展，可以考虑用新的方法来实现更快并且更准确的乐器选材，根据查阅文献发现将深度学习用于木材振动声音信号的识别方面的研究非常少，所以尝试将深度学习卷积神经网络方法应用到木材振动信号分类识别方面。

深度学习[8]是基于人工神经网络发展起来的技术，随着深度学习的快速发展，声音识别领域取得的成绩也有了突破性的进展，其中卷积神经网络 CNN[9]已经成为图像识别和声音识别中的热门研究方向之一，它拥有着独特的网络结构，通过引入“卷积”和“降采样”操作，可以实现对多维的输入特征进行处理，其实卷积神经网络早在 2006 年以前就被人们提出来了，但此时的 CNN 在小图片的识别方面效果较好，不适合识别大规模的数据。2012 年 Hinton 教授[10]和他的学生利用更深层次的卷积神经网络模型在闻名世界的 Image Net 问题上，结果取得了出乎意料的好成果，这标志着 CNN 在图像识别领域逐渐占据了主导地位。在 2014 年左右，来自南洋理工大学的 Dennis 博士[11, 12]提出了声音的频谱图像特征，他是受启发于 Zue 的“频谱阅读”(spectrogram reading)[13]，并取得了很好的识别结果，这标志着声音事件识别的突破性进展。多伦多大学和 IBM 沃森研究组[14]对于 CNN 模型的声音输入特征分类从而测试识别效果，最后得出将声音特征滤波器组系数（filter bank）作为声音特征输入时的识别效果最好。Hamid 等人[15]将 NN/HMM 与 CNN 结合用于语音识别中，在 LVCSR 数据库上取得了成功。

因此，基于卷积神经网络的图像识别[16]技术、声音识别技术得到了广泛的关注，为当前图像以及声音识别领域的研究热点之一。对木材样本的敲击声是一种典型的瞬态声，木材是具有弹性的固体材料，能依靠它的弹性介质作用来传递声波的能量。木材能够在冲击力的作用下，由自身的振动辐射声能，从而发出优美音色的乐音，其声学性能越好，具有的声共振特性就越优良，这种特性是木材能够广泛应用于乐器制作的重要依据[17]。这种声音分类识别的关键在于特征提取，而声音信号的特征之一语谱图[18,19]具有很强的实用价值，它给出了木材振动声音的多种特征信息，动态的显示出信号频谱的变化情况，充分反映了声音信号的频谱特性和完整信息。所以本文提出的通过提取木材振动声音信号的语谱图特征并应用卷积神经网络识别正是满足此需求的新方法。

1　卷积神经网络的结构

卷积神经网络是一种经典的前馈神经网络，包括正向传播和反向传播两个过程。卷积神经网络一般包括卷积层（convolution layer）、下采样层（pooling layer）和全连接层（fully-connection layer），图 1 是一种比较经典的卷积神经网络 LetNet-5[20]网络结构。

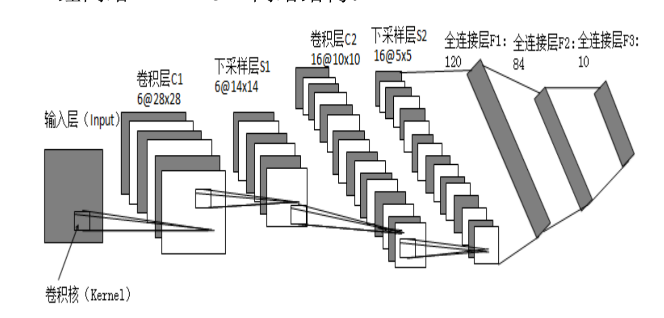


图 1　LeNet-5 网络结构

1.1　卷积层

卷积层（convolutional layer）也被称为 Conv.layer，它构成了 CNN 的基础，由多个特征面（Feature Map）组成，直接对原始输入信号（一般为二维信号，图像）进行卷积操作。卷积核是一个权值矩阵，它的大小自行设置，通常选择 3×3 或 5×5 模板。卷积核以一定的步长在特征图上进行“滑动”，每滑动一次就进行一次卷积操作，通过多次的卷积操作提取输入信号的不同特征，每一个卷积核能提取一种特征，这样 n 个卷积核就可以提取 n 种特征。通常卷积层的计算形式如式（1）所示。

$$x_j^l = f\sum_{i \in M_j}(x_j^{l-1}k_{ij}^l + b_j^l) \tag{1}$$

其中：$f(\cdot)$ 表激活函数（可以是 Sigmoid、Tanh 等非线性函数），K 代表卷积核，l 代表卷积层数，M_j 是输入层的感受野，b 代表每个输入图的一个偏置值。

1.2　池化层

卷积层是池化层的输入层，池化层即图像的下采样层，通常会在连续的卷积层之间定期插入一个池化层，它的功能是减少网络中的计算量和参数，从而也可以控制过拟合。本文会使用池化函数来进一步调整这一层的输出，池化函数使用某一位置的相邻输出的总体统计特征来代替网络在该位置的输出。例如，本文中用到的最大池化函数（max　pooling）给出相邻矩形区域内的最大值。其他常用的池化函数包括相邻矩形区域内的平均值、L2 范数以及基于距中心像素距离的加权平均函数。池化层的一般形式如式（2）所示。

$$x_j^l = f(\beta_j^l \mathrm{p}(x_j^{l-1}) + b_j^l) \tag{2}$$

其中：β 表示权重系数，p（·）表示池化函数。

1.3　全连接层

全连接层的每个神经元将和上一层的全部神经元进行全连接，经多个卷积层和池化层后，连接着一个或多个全连接层来输出最后的分类结果。最后一层的输出值传递到输出层，对于

本文的多分类问题，采用 softmax 层可以得到当前样本属于不同种类的概率分布情况。

1.4　其他常用层

在 CNN 网络结构中还有一些常用的层，如激活层和 Dropout 层等。一般来说，如果训练数据集不是很多时，在训练过程中可能会学习到训练数据的噪声，导致出现过拟合现象，它的具体表现就是模型在训练集上和测试集上的效果相差较大，模型泛化能力弱，所以本文在网络的训练部分加入这 2 层可以有效的防止过拟合。

本文的激活层采用修正线性单元（rectified linear unit, ReLU）激活函数[22]，它能把输入值“压缩”到 0 到正无穷大范围内，当输入值为正数值时，ReLU 函数将它直接传递；当输入值为负数值时，ReLU 函数将它设置为零，除此之外还有 Tanh 和 Sigmoid 函数等。相比之下，激活函数 ReLu 的收敛速度快很多，会加快网络的训练速度，只需一个阈值就可以得到激活值，不需要去计算大量复杂的运算。ReLU 激活函数是一个分段函数，其数学形式如式（3）所示。

$$\mathrm{ReLU}(x)=\max(0,x) \tag{3}$$

很显然，从公式可以看出，输入信号小于 0 时，输出都是 0；输入信号大于 0 的情况下，输出等于输入。由此可以看出 ReLu 函数会使部分神经元的输出结果为 0，不仅使网络具有了稀疏性，还减少了参数间的依存关系，有效的缓解了过拟合问题。

Dropout[21]层通常放在池化层和全连接层后，是一种很好的防止过拟合方法，本文在网络的训练部分加入 Dropout 技术，相当于从原来的网络变成一个更稀疏的网络继续训练，每一个节点随机以一定的概率 p 被设置为零，若实验中将 p 设置为 0.5，表示随机输出 50%的神经元，由于网络节点不会对其他节点的即时状态作出响应，阻止了某些节点仅仅在其他特定节点下发挥作用的情况，稀疏后的网络由所有保留的单元组成。一个具有 n 个单元的神经网络，可被看做是 2^n 个可能的稀疏神经网络的集合。这些网络共享权重，所以参数的总数仍然是 O(n^2)，或者更少。对于每个训练模型的每一个演示，一个新的稀疏网络被抽样并训练。所以训练一个使用 dropout 技术的神经网络可以被看做是训练一个具有广泛的权值共享的 2^n 的细化网络的集合。网络的泛化能力得到提高，具有更好的适应性。

2　特征提取

目前在声音识别领域常用的声音特征参数有线性预测倒谱系数(LPCC)、Mel 频率倒谱系数(MFCC)，其中 MFCC 特征是基于符合人耳听觉特性提出的，它同时结合人耳的听觉机理与声音的产生机制特性，不依赖于全极点声音产生模型的假定，与声音信号的实际频率成非线性对应关系，在一定程度上模拟了人耳对声音的处理特点，在噪声环境中表现出不错的鲁棒性，近年来被广泛的应用于各类型声学中，并取得了良好的效果。而 LPCC 是基于发音模型建立的，此参数没有充分考虑人耳的听觉特性，在噪声环境中的鲁棒性不好，尽管 MFCC 与 LPCC 特征相比有明显优势，但是基于本文的算法考虑，声音信号的语谱图特征要更有优势，语谱图不仅可以保留更多的信息（包括可能的冗余信息），而且还能够使用卷积和池化操作来表示和处理一些典型的声音不变性和变异性。因此本文重点研究提取语谱图特征，语谱图即声音频谱图，即将音频转换为语谱图图像，不仅可以使用语音信号处理的知识，还可以融合图像处理技术，即将图像处理技术应用到语音处理方面，因此把声谱图应用于声音事件识别中是很有前景的。语谱图显示了大量的和声音相关的特征信息，它的横坐标表示时间，纵坐标表示频率，其坐标（x,y）对应的点表示在时间 x，频率 y 上的语音数据能量（即声音强度），能量是通过颜色的浓淡来表示出来的，颜色越深，表示该点的语音能量越强，这样就采用了二维平面来表达三维信息。语谱图有很强的实用价值，其综合了时域波形和频谱图的特征，明显的显示出声音特征随时间变化的情况，这在波形图中是无法展现的，如图 2 所示。

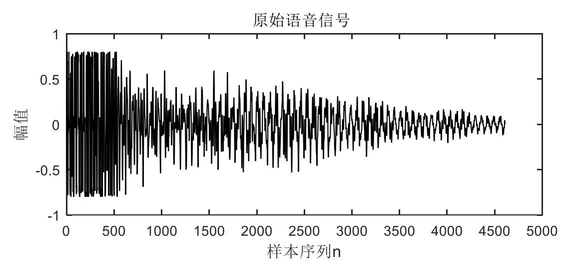


图 2　原始木材振动信号的波形图

语谱图的整个提取过程如图 3 所示。

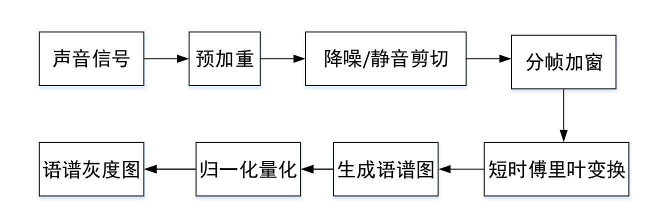
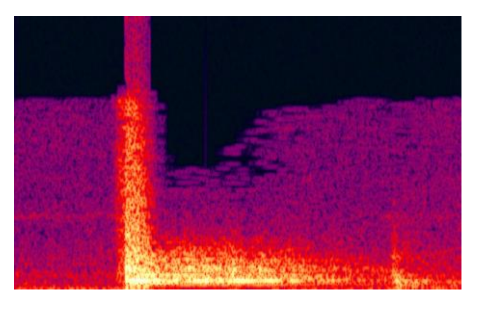


图 3　特征提取算法流程图

声音信号中包含木材振动声音之外的噪声和静音段，这些因素都会影响声音特征的性能，噪声会减弱信号中的部分有效信息，而静音段会影响声音信号在语谱图的位置，因此为使实验结果更加准确，需要对声音进行降噪和静音剪切，原始语谱图和经过处理后的语谱图如图 4 和 5 所示。


图 4　原始木材振动信号语谱图

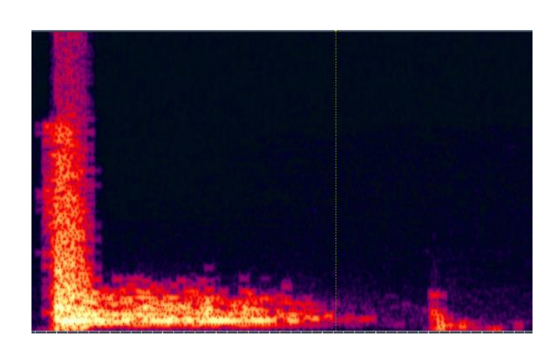


图 5　经过去噪、静音剪切后的语谱图

首先对声音信号进行预加重处理，目的是提升木材振动声音的高频部分，更有助于进行整个振动声音信号的频谱分析，方法是采用一个一阶的高通滤波器，其数学表达式如下：

$$H(z)=1-\mu Z^{-1} \tag{4}$$

其中：μ 为预加重系数，取值接近于 1，本实验中 μ=0.97。

然后进行分帧加窗和短时傅里叶变换处理，即

$$X(n,k)=\sum_{m=0}^{M-1} x_n(m)\omega(m)e^{-j2\pi km} \quad 0\le k\le N-1 \tag{5}$$

其中：n 是时域采样点序列，n=0,1，⋯，N-1（N 是信号长度）；$x_n(m)$ 为经过分帧处理后的声音信号，$m=0,1,\cdots,M$-1，其中 m 是帧同步时间序号，n 是帧序号；$\omega(m)$ 为汉明窗函数，可以减轻由加窗操作导致的声音不连续性，其定义为

$$\omega(\text{n})=\begin{cases}0.54-0.46\cos(\dfrac{2\pi n}{L-1}) & 0\le \text{n}\le L-1\\ 0 & \text{other}\end{cases} \tag{6}$$

其中：L 为窗长，一般情况下帧长取 10~30ms 时可认为信号是平稳的，本实验取帧长为 512 点，帧与帧之间的交叠部分为帧移，帧移一般取帧长的一半，即 256 点；由上述过程可得到语谱图 $\hat{X}$。

其次，采用对数能量的方法生成语谱图 $S_{Log}(x,y)$，即

$$S_{Log}(x,y)=20\log g(|\hat{X}(x,y)|) \tag{7}$$

其中：x\{1，2，...,X\}，y \{1，2，...,Y\}为语谱图像素的二维坐标，其中 X，Y 分别表示语谱图横、纵坐标的最大值。

最后，利用最大最小归一化（mapminmax）方法对语谱图进行归一化，归一化的作用是为了实现数据规范化，使其灰度变化范围为 [0,1]，从而保证样本数据具有统一的统计分布性，以便后面的处理更加准确和方便，归一化过程定义如下：

$$G(x,y)=\frac{S(x,y)-\min(S(x,y))}{\max(S(x,y))-\min(S(x,y))} \tag{8}$$

其中：min(S（x，y）)，max(S（x，y）)分别代表时频矩阵中最小值和最大值。通过归一化处理得到[0,1]内的时频矩阵 G（x，y）。然后以 G（x，y）作为灰度强度值得到灰度语谱图。

提取木材振动信号的特征来进一步分析和识别木材的质量是非常关键的一步，它直接影响后面实验的识别速度和准确度，本文使用 STFT 把声音信号转换成语谱图以提取特征，并把语谱图的空白部分剪切掉以减少计算量，提高精度的匹配。

3　实验方法与结果分析

3.1　实验声音样本集及实验环境

本文实验声音数据是通过在安静的实验室里利用对相同大小、不同优劣质量的的兰考泡桐木板进行敲击的方式获取的，采集设备为录音笔，通过对 3 种不同优劣等级的木板的不同位置进行反复声波采集，得到采样率为 44.1KHz 的单声道 16 位 wav 格式的声音文件，然后使用 CoolEdit 软件对原始声音样本进行切割筛选，进而得到单个的声音样本以方便后续实验，测试样本格式是 116*80 像素的 1 通道，灰度图，数量为 7319 个样本。测试集是 2196 个样本。实验数据的具体信息描述如表 1 所示。

表 1　实验数据具体信息

材料类别	板材数量	训练样本数/个	测试样本数/个	总样本数/个
优良	11	1785	765	2550
一般	10	1636	701	2337
较差	10	1702	730	2432

最新的神经网络库 keras[23]引起了广泛的关注。用 Theano 或 TensorFlow 作为后端，本文对原始木材振动信号进行特征提取时，采用 matlab2016a 编程软件，搭建的卷积神经网络均基于 Linux 平台，Keras 运行在 Python3.5 后端的 Tensorflow 框架。

3.2　实验及结果分析

3.2.1 基于网格搜索的 CNN 模型调整

本文数据量不是很大，所以考虑使用基于网格搜索[24]的方法调整不同参数设置来拟合 CNN 模型，从而避免参数选择的随意性和盲目性。本文对优化器 optimizer、Dropout 率取值、迭代次数 epochs、批量 batch_size 等参数通过网格搜索的办法进行参数优化，实验参数设置及结果如表 2 所示，根据实验结果得出以下结论：

a）批量太小（比如 1）阻碍了网络的融合，而批量太大缩小了迭代次数，从而导致需要更长的时间去达到很好的精度，根据程序结果建议将批量值设置为 64。

b）正则化，Dropout 是一种简单并常用的正则化技术，适用于防止过度拟合。dropout 率经过测试从 0.5 到 0.2，并通过测试建议取值为 0.3。

c)迭代次数是指将训练集输入到神经网络中进行训练的次数。当测试错误率和训练错误率相差较小时，当前的迭代次数被认为是最合适的，否则需调整网络结构或增大迭代次数，本文选择 10 次作为 epochs 最佳取值。

d）优化器 optimizer 的种类很多，比如 SGD、RMSprop、Adadelta、Adam，经过实验程序测试，选择 Adam 优化器进行后续实验。

表 2　网格搜索结果

类别	Dropout	Batch_size	Epochs	Optimizer
1	0.2	32	5	sgd
2	0.3	64	10	adam
3	0.4	128	15	rmsprop
4	0.5	256	20	adadelta
最优解	0.3	64	10	adam

3.2.2 不同 CNN 网络结构对比实验

本文搭建的 CNN 采用 3 类结构，参数设置如下所示（卷积层用 C 表示，降采样层用 P 表示，全连接层用 F 表示），卷积层后均加入激活层（RuLU）,池化层后均加入 Dropout 层（Dropout 率取值为 0.3）：

a）采用“C1+P1+F1+F2 层”结构，C1 层卷积核设置为 16 个，大小为 3*3；P1 大小为 2*2，采用最大值池化输出；F1 层节点数为 64，F2 层节点数就是输出类别即 3，表示为 CNN-1。

b）采用“C1+P1+C2+C3+P2+F1+F2 层”结构，其他层与 CNN-1 结构相同，C2 层卷积核设置为 32 个，大小为 3*3；C3 层卷积核设置为 32 个，大小为 5*5；P2 大小为 2*2，采用最大值池化输出；表示为 CNN-2。

c）采用“C1+C2+P1+C3+C4+P2+F1+F2 层”结构，其他层与 CNN-1 结构相同，C2 层卷积核设置为 16 个，大小为 3*3；C3 层卷积核设置为 32 个，大小为 3*3；C4 层卷积核设置为 32 个，大小为 5*5；P2 大小为 2*2，采用最大值池化输出；表示为 CNN-3。

对三种卷积结构进行训练和分类效果测试，实验结果如图 6~11 所示。

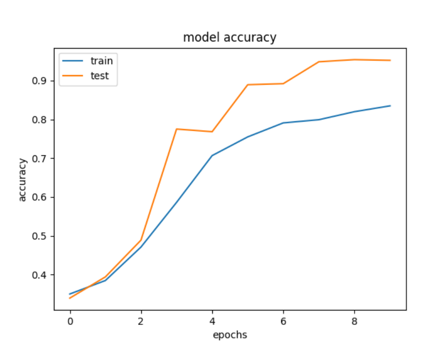


图 6　CNN-1 实验准确率

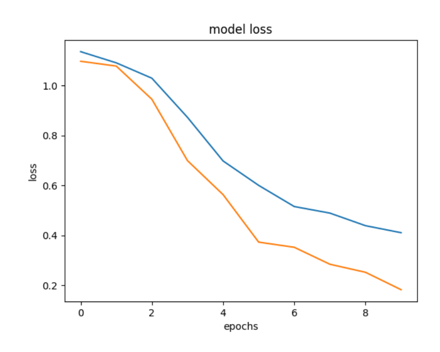


图 7　CNN-1 实验损失值

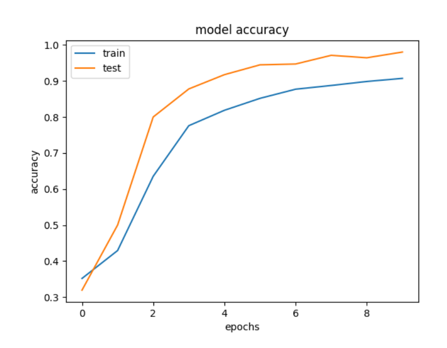


图 8　CNN-2 实验准确率

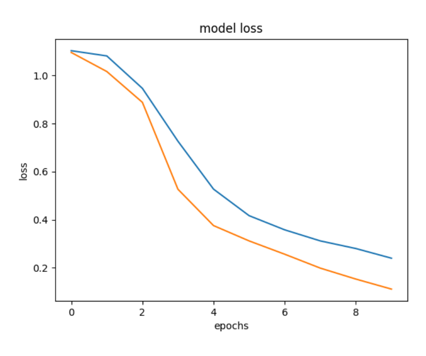


图 9　CNN-2 实验损失值

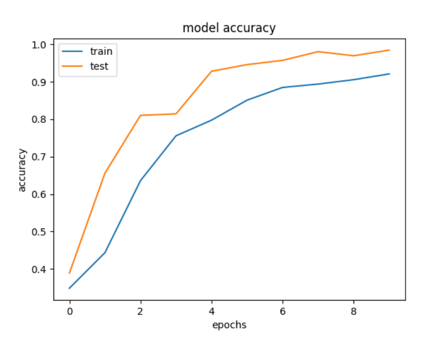


图 10　CNN-3 实验准确率

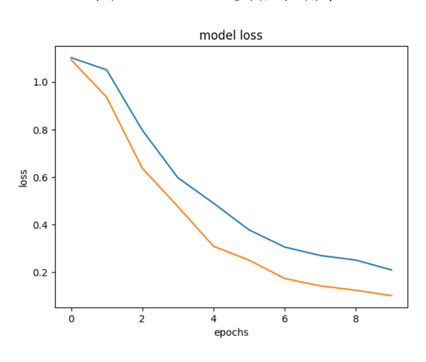


图 11　CNN-3 实验损失值

由以上结果图可知，采用本文提出的 CNN-2 分类的识别率较高，并且准确率曲线和损失值曲线最平滑，实验运行时间适中；而 CNN-1 尽管实验用时最少，测试集识别率较高，但训练集识别率偏低；CNN-3 的曲线图有些不平滑，而且识别率和 CNN-1、CNN-2 相比较低，并且运行时间最长。一次 epoch 是一次完整迭代(所有样本都训练过)，这里本文用了 10 次迭代，上图中的 loss 表示训练集损失值，acc 表示训练集准确率；val_loss 表示测试集上的损失值，val_acc 表示测试集上的准确率。在 CNN-2 模型中，最后一次迭代就可以收敛到 96%的预测

准确率了，对于每个类别的木材振动声音识别准确率和召回率等参数结果如表 3 所示。

表 3　CNN-2 实验结果

类别	准确率	召回率	F1-值	测试样本数量
0（优良）	0.95	0.97	0.95	765
1（一般）	0.96	0.97	0.96	701
2（较差）	0.97	0.94	0.97	730
平均率	0.96	0.96	0.96	2196

综合上述对比实验结果可得，CNN-2 结构取得的实验效果最好，并且层数越少的结构花费的运行时间越少，但是越少的层数会伴随着越多的权重，这同时也增加了记忆负担，这也是 CNN-2 优于 CNN-1 的原因。一个卷积核只能检测一种特征（比如图片的纹理，垂直方向的边缘），而语谱图中的特征往往比较复杂，一个卷积核显然是不够的，所以在神经网络的卷积层中会有多个卷积核，不同的特征图含有不同的特征，这样卷积层的输出就会有多层。而卷积核数目的设置一般按照偶数倍递增，特征图越多是为了提取更多的特征，但同时卷积核越多，需要处理的参数就越多，为降低模型的复杂度，所以本文中卷积核的个数设置均没有超过 32 个。其中 CNN-3 网络结构更加复杂，与 CNN-2 相比多一层带有 32 个卷积核的卷积层，因此训练 CNN-3 结构需要更多的参数，训练效率较低，而且正确率也没有得到提升，这说明在卷积神经网络的训练过程中，简单的增加神经网络中的卷积核个数，提高网络结构的复杂性并不能对应的提高其分类性能。而其他因素如好的网络经验参数、合适的迭代次数、批量值等对分类器的性能也能产生一定的影响。

语谱图能有效的表示不同优劣质量木材的振动声音信号的特征，相比于其他网络，CNN 的降采样操作具有尺度平移不变性，降低了特征图的维度，可以避免过拟合；CNN 局部区域感知操作，既可以降低噪声，又可以增强语音特征，而且能更好的分析能量分布情况。

4　结束语

本文在对木材振动信号识别问题上引入了深度学习卷积神经网络方法，并结合图像识别技术提出提取木材振动信号的语谱图特征。本文还在参数调优以及卷积神经网络结构上展开研究，将 CNN 作为分类器应用到识别系统中，通过网格搜索技术获得最优参数，实验结果表明，所提出的算法具有较好的有效性，和之前的检测方法相比，更具有科学性和实用性。本文实验中的样本数量以及网络性能方面还存在不足，在未来的研究工作中，我会收集更多的兰考泡桐的声音振动信号数据样本，并尝试进一步改进网络结构来提高识别的准确率。

参考文献：

[1] 罗嘉琪，唐衡．中国民族乐器传承与发展 [J]．前沿, 2013, 30 (7): 166-167.

[2] 李源哲，李先泽，汪溪泉，等．几种木材声学性质的测定 [J]．林业科学, 1962, 7 (1): 59-66.

[3] 沈隽，刘一星，刘振波，等．纤丝角对云杉属木材声振动特性的影响 [J]．东北林业大学学报, 2002, 30 (5): 50-52.

[4] Ono T, Norimoto M. Study on Young's modulus and internal friction of wood in relation to the evaluation of wood for musical instruments [J]. Limnology & Oceanography, 1983, 22 (4): 611-614.

[5] Tonosaki M, Okano T, Asano I. Vibrational properties of Sitka Spruce with longitudinal vibration and flexural vibration [J]. Mokuzai Gakkaishi, 1983.

[6] Sobue N. Measurement of Young's modulus by the transient longitudinal vibration of wooden beams using a fast Fourier transformation spectrum analyzer [J]. Journal of the Japan Wood Researh Society, 1986, 32 (9): 744-747.

[7] Treu A, Hapla F. Study on the quality of spruce and fir resonance wood [J]. Allgemeine Forst-und Jagdzeitung, 2000, 171 (12): 215-222.

[8] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521 (7553): 436-444.

[9] 周飞燕，金林鹏，董军．卷积神经网络研究综述 [J]．计算机学报, 2017, 40 (06): 1229-1251.

[10] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural network [J]. Science, 2006, 313 (5786): 504-507

[11] Dennis J W. Sound event recognition in unstructured environments using spectrogram image processing [D]. Singapore: Nanyang Technological University, 2014.

[12] Dennis J, Tran H D. Enhanced local feature approach for overlapping sound event recognition [C]// Proc of Summit and Conference on Asia-Pacific Signal and Information Processing Association. 2015: 1-4.

[13] Zue V. Notes on speech spectrogram reading [C]// MIT Lecture Notes. Cambridge, MA, 1991.

[14] Sainath T N, Mohamed A R, Kingsbury B, et al. Deep convolutional neural networks for LVCSR [C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2013: 8614-8618.

[15] Abdel-Hamid O, Mohamed A R, Jiang H, et al. Convolutional neural networks for speech recognition [J]. IEEE//ACM Trans on Audio Speech & Language Processing, 2014, 22 (10): 1533-1545.

[16] 邓柳，汪子杰．基于深度卷积神经网络的车型识别研究 [J]．计算机应用研究, 2016, 33 (03): 930-932.

[17] 李坚．木材科学 [M]. 3 版．北京：科学出版社, 2014.

[18] 梁士利，魏莹，潘迪，等．基于语谱图行投影的特定人二字汉语词汇识别 [J]．吉林大学学报：工学版, 2017, 47 (1): 294-300.

[19] 陶华伟，查诚，梁瑞宇，等．面向语音情感识别的语谱图特征提取算法 [J]．东南大学学报：自然科学版, 2015, 45 (05): 817-821.

[20] LeCun Y. LeNet-5, convolutional neural networks [EB/OL]. 2015. http://yann. lecun. com/exdb/lenet.

[21] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning

Research, 2014, 15 (1): 1929-1958.

[22] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3 (4): págs. 212-223.

[23] Francois C. https://keras. io/ [EB/OL].

[24] 李瀚，杨晓峰，邓红霞，等. 基于网格搜索算法的 PCNN 模型参数自适应 [J]. 计算机工程与设计, 2017, 38 (1): 192-197.